

Preuve de Cohérence

Une couche de gouvernance et de vérification pour les maillages IA fédérés

Sylvain Cormier
 Paraxiom Research
 research@paraxiom.io

Janvier 2026

Énoncé du problème

Les systèmes d'IA fédérés et distribués reposent sur des nœuds non fiables et hétérogènes qui contribuent des mises à jour de modèles, des sorties d'inférence ou du travail d'entraînement local. Aujourd'hui, les critères d'acceptation sont faibles : agrégation statistique, confiance envers les opérateurs ou détection grossière d'anomalies.

Cela crée quatre risques :

1. **Contributions hallucinées ou incohérentes** acceptées dans les modèles partagés
2. **Déstabilisation silencieuse** des modèles globaux par des nœuds malveillants
3. **Aucune notion objective de « travail utile »** dans les maillages IA
4. **Gaspillage d'énergie** sur la vérification redondante par recalcul

Il n'existe actuellement aucun équivalent de « preuve de travail utile » pour l'IA distribuée.

Idée centrale

La Preuve de Cohérence est une couche de vérification légère qui prouve qu'une contribution IA était cohérente en interne, non divergente et sûre à accepter dans un système IA partagé—**sans faire confiance au nœud** qui l'a produite et **sans réexécuter le modèle**.

Nous ne prouvons pas qu'une IA a raison—nous prouvons qu'elle est restée cohérente.

Ce que « Cohérence » signifie (opérationnellement)

Chaque contribution IA produit, en plus de sa sortie, une **empreinte de cohérence** :

Signal	Ce qu'il mesure	Calcul
Entropie d'attention	Stabilité de distribution entre les têtes	$O(n)$ par couche
CV spectral	Coefficient de variation des valeurs propres	$O(n^2)$ une fois
Consistance topologique	Déviation des voisnages attendus	$O(n)$ par couche
Taux de dérive	Changement des normes d'état caché	$O(1)$ par couche

Ces signaux sont **conscients de l'architecture** (calibrés par famille de modèle), **peu coûteux à calculer** ($\sim 5\%$ de surcharge) et **déterministes à vérifier**.

Intégration dans l'apprentissage fédéré

Flux existant : Nœud → Entraînement/Inférence local → Mise à jour envoyée → Agrégation

Avec Preuve de Cohérence : Nœud → Entraînement/Inférence local → Empreinte de cohérence (5% surcharge) → Porte de vérification (0,1% surcharge) → Accepté/Rejeté → Agrégation

C'est une **couche de gouvernance parallèle**, pas une réécriture. Aucun changement aux algorithmes, architectures ou outils d'orchestration.

Pourquoi c'est important

1. Sensible à l'architecture

Différents modèles ont différentes enveloppes de stabilité. Notre recherche le démontre empiriquement :

Modèle	Effet de la contrainte toroïdale
Phi-2 (2,78B)	50% de réduction des hallucinations
TinyLlama (1,1B)	180% d' augmentation des hallucinations

Même contrainte, effet opposé. Les corrections universelles n'existent pas. La gouvernance doit être consciente de l'architecture.

2. Conscient des hallucinations

Les hallucinations sont traitées comme des **échecs de cohérence**, pas seulement des erreurs factuelles. Un modèle qui dérive hors de son enveloppe est signalé avant que sa sortie ne se propage.

3. Résistant aux adversaires

Les empreintes de cohérence sont dérivées des dynamiques internes du modèle. Falsifier la cohérence est computationnellement équivalent à faire un travail cohérent.

4. Économe en énergie

Approche	Coût énergétique
Recalcul complet	100%
Vérification Preuve de Cohérence	↓ 1%

Pour l'IA fédérée à grande échelle : **réduction de 99%+** de l'énergie de vérification. C'est une gouvernance IA verte : la confiance par les mathématiques, pas par la force brute.

Preuve de travail utile (reformulée)

Les systèmes traditionnels prouvent l'énergie brûlée (PoW) ou le capital verrouillé (PoS).

La Preuve de Cohérence prouve : **Ce travail IA était utile car il a préservé ou amélioré la stabilité du système.**

Statut actuel

- Implémenté et testé sur de vrais LLMs
- Découverte d'effets dépendants de l'architecture (critique pour la gouvernance)
- Publié : DOI: [10.5281/zenodo.18267913](https://doi.org/10.5281/zenodo.18267913)
- Code ouvert : github.com/Paraxiom/topological-coherence

Ce que cela permet

Capacité	Bénéfice
Primitive de gouvernance	Accepter/rejeter les contributions objectivement
Confiance sans contrôle central	Les noeuds prouvent leur propre cohérence
Efficacité énergétique	Vérifier sans recalculer
Scalabilité	Vérification O(1) par contribution

Résumé en une phrase

La Preuve de Cohérence est une couche de vérification qui prouve que le travail IA distribué était stable et utile—pas seulement calculé—avant qu'il ne soit accepté par le système, à 99% moins de coût énergétique que le recalcul.